METHOD OF AND SYSTEM FOR PROVIDING ADAPTIVE RESPONDENT TRAINING IN A SPEECH RECOGNITION APPLICATION

Cross References To Related Applications

5   This application claims the benefit of priority from commonly owned U.S. Provisional Patent Application Serial Number 60241,757, filed October 16, 2000, entitled ADAPTIVE USER TRAINING FOR SPEECH RECOGNITION APPLICATION.

Field of the Invention

10   The present invention relates generally to a method of and system for providing adaptive respondent training in a speech recognition algorithm, and more particularly to a method of and system for determining the level of understanding and capability of a respondent to a telephonic speech recognition application, and both providing specific instructions to the respondent regarding the application and adapting the application to suit 15   the capabilities of the respondent.

Background of the Invention

       In the new, connected economy, it has become increasingly important for companies or service providers to become more in tune with their clients and customers. 20   Such contact can be facilitated with automated telephonic transaction systems, in which interactively-generated prompts are played in the context of a telephone transaction, and the replies of a human user are recognized by an automatic speech recognition system. The answers given by the respondent are processed by the system in order to convert the spoken words to meaning, which can then be utilized interactively, or stored in a database. 25   In order for a computer system to recognize the words that are spoken and convert these words to text, the system must be programmed to phonetically break down the words and convert portions of the words to their textural equivalents. Such a conversion requires an understanding of the components of speech and the formation of the spoken word. The production of speech generates a complex series of rapidly changing acoustic pressure 30   waveforms. These waveforms comprise the basic building blocks of speech, known as

phonemes. Vowel and consonant sounds are made up of phonemes and have many different characteristics, depending on which components of human speech are used. The position of a phoneme in a word has a significant effect on the ultimate sound generated. A spoken word can have several meanings, depending on how it is said. Speech scientists have identified allophones as acoustic variants of phonemes and use them to more explicitly define how a particular word is formed.

While there are several distinct methods for analyzing the spoken word and extracting the information necessary to enable the recognition system to convert the speech to word-strings, including Hidden Markov modeling and neural networks, these methods generally perform similar operations. The differences in these methods are typically in the manner in which the system determines how to break the phonetic signal into portions that define phonemes. Generally, a speech recognition system first converts an incoming analog voice signal into a digital signal. The second step is called feature extraction, wherein the system analyzes the digital signal to identify the acoustic properties of the digitized signal. Feature extraction generally breaks the voice down into its individual sound components. Conventional techniques for performing feature extraction include subband coding Fast Fourier Transforms and Linear Predictive Coding. Once the signal has been analyzed, the system then determines where distinct acoustic regions occur. The goal of this step is to divide the acoustic signal into regions that will be identified as phonemes which can be converted to a textual format. In isolated word systems, this process is simplified, because there is a pause after each word. In continuous speech systems, however, this process is much more difficult, since there typically are no breaks between words in the acoustic stream. Accordingly, the system must be able not only to break the words themselves into distinct acoustic regions, but must also be able to separate consecutive words in the stream. It is in this step that conventional methods such as Hidden Markov modeling and neural networks are used. The final step involves comparing a specific acoustic region, as determined in the previous step, to a known set of templates in a database in order to determine the word or word portion represented by the acoustic signal region. If a match is found, the resulting textual word is output from the system. If one is not, the signal can either be dynamically manipulated in order to increase

the chances of finding a match, or the data can be discarded and the system prompted to repeat the query to the respondent, if the associated answer cannot be determined due to the loss of the data.

In customer service applications, it is important for service providers to be able to obtain information from, or to provide information to, their customers. Oftentimes, service providers will need to contact customers via the telephone to obtain or provide the desired information. In order to reduce the costs associated with such information exchanges, many service providers utilize automated telephone calling devices to contact customers. While the automated telephone calling devices are extremely capable of converting spoken words into text phrases and thereby obtaining valuable information from respondents, in some cases, the respondents are not capable of providing adequate responses to the posed questions, or do not understand what is involved in an automated telephonic application. Prior art speech recognition applications are not able to identify that the respondent is having trouble with the application and then adjust the application accordingly. This results in wasted time and money for the company in charge of the survey and in frustration on the part of the respondent.

Summary of the Invention

The present invention is directed to a method for adaptive training of a respondent to a telephonic speech recognition application. The method is used in connection with the speech recognition application to enable the administrator of the application to explain the function of the application, to train the respondent in how to effectively respond to the queries in the application and to adapt the application to the needs of the respondent, based on the initial responses given by the respondent.

According to one aspect of the invention, a method of conducting a telephonic speech recognition application is disclosed, including:

    A. making telephonic contact with a respondent;

    B. presenting the respondent with at least one introductory prompt to reply to;

    C. utilizing a speech recognition algorithm to process the audio responses given by the respondent to determine a level of capability of the respondent;

D. based on the audio responses, presenting the respondent with one of:

at least one prompt associated with an application; and

an explanation of the operation of the speech recognition application.

The explanation may include at least one of a sample prompt and instructions on

5 responding to the at least one prompt of the application.

According to another aspect of the invention, a system for conducting a telephonic speech recognition application is disclosed, including:

an automated telephone device for making telephonic contact with a respondent; and

10 a speech recognition device which, upon the telephonic contact being made, presents the respondent with at least one introductory prompt for the respondent to reply to; receives a spoken response from the respondent; and performs a speech recognition analysis on the spoken response to determine a capability of the respondent to complete the application;

15 wherein, if the speech recognition device, based on the spoken response to the introductory prompt, determines that the respondent is capable of competing the application, the speech recognition device presents at least one application prompt to the respondent; and

wherein, if the speech recognition device, based on the spoken response to the

20 introductory prompt, determines that the respondent is not capable of completing the application, the speech recognition system presents instructions on completing the application to the respondent.

Brief Description of the Drawings

25 The foregoing and other objects of this invention, the various features thereof, as well as the invention itself may be more fully understood from the following description when read together with the accompanying drawings in which:

Fig. 1 is a schematic block diagram of the system for providing adaptive respondent training in accordance with the present invention;

Fig. 2 is a flow diagram of a method for providing adaptive respondent training in accordance with the present invention; and

Figs. 3A-3C are flow diagrams showing an example of the instruction stage of the present invention.

5

Detailed Description

As set forth above, many customer-oriented organizations, including retail operations, service organizations, health care organizations, etc. rely on interactions with their customers in order to obtain valuable information that will enable the organizations to

10 optimize their operations and to provide better service to the customers. Telephonic speech recognition applications, in which specific prompts about the organization's products or services, ' enable the organizations to obtain information from customers ' in a manner which consumes very little time and which does not require repeat visits to the organization's location. For many organizations, these types of interactions are much less

15 troublesome for customers who might have difficulties in traveling.

While speech recognition applications can be an extremely efficient way to gather information from respondents, if the respondent is not able to respond to the prompts of the survey or does not understand the survey process or how to respond to certain types of queries, the process can be frustrating for respondent, thus inhibiting future interactions

20 with the respondent, and the process can be costly and time consuming for the organization providing the service.

The present invention includes a method and system for determining whether a respondent is capable of responding to the prompts in a telephonic speech recognition application and what extra explanations or instructions, with modified application

25 functionality, might be required to assist the respondent in completing the application. The method is incorporated into the application, and responses to introductory prompts of the application direct the application to present prompts to the respondent that will enable the respondent to learn how to correctly complete the application.

Referring now to Figs. 1-3, a preferred embodiment of the present invention will be

30 described. System 12, Fig. 1, includes an automated telephone calling system 14 and a

speech recognition system 16. Preferably, the automated telephone calling system 14 is a personal computer such as an IBM PC or IBM PC compatible system or an APPLE MacINTOSH system or a more advanced computer system such as an Alpha-based computer system available from Compaq Computer Corporation or SPARC Station

5   computer system available from SUN Microsystems Corporation, although a main frame computer system can also be used. In such a system, all of the components of the system will reside on the computer system, thus enabling the system to independently process data received from a respondent in the manner described below. Alternatively, the components may be included in different systems that have access to each other via a LAN or similar

10   network. For example, the automated telephone calling device 14 may reside on a server system which receives the audio response from a telephone 18 and transmits the response to the speech recognition device 16.

The automated telephone calling system 14 may also include a network interface that facilitates receipt of audio information by any of a variety of a networks, such as

15   telephone networks, cellular telephone networks, the Web, Internet, local area networks (LANs), wide area networks (WANs), private networks, virtual private networks (VPNs), intranets, extranets, wireless networks, and the like, or some combination thereof. The system 10 may be accessible by any one or more of a variety of input devices capable of communicating audio information. Such devices may include, but are

20   not limited to, a standard telephone or cellular telephone 18.   Automated telephone calling system 14 includes a database of persons to whom the system 12 is capable of initiating or receiving telephone calls, referred to hereinafter as the "target person", a telephone number associated with each person and a recorded data file that includes the target person's name. Such automated telephone calling devices are known in the art.

25   As is described below, the automated telephone calling system 14 is capable of initiating or receiving a telephone call to or from a target person and playing a prerecorded greeting prompt asking for the target person. The system 14 then interacts with speech recognition system 16 to analyze responses received from the person on telephone 18.

Speech recognition system 16 is an automated system on which a speech

30   recognition application, including a series of acoustic outputs called prompts, which

comprise queries about a particular topic, are programmed so that they can be presented to a respondent, preferably by means of a telephonic interaction between the querying party and the respondent. However, a speech recognition application may be any interactive application that collects, provides, and/or shares information. As examples, in the

5   present invention, a speech application may be any of a group of interactive applications, including consumer service or survey applications; Web access applications; customer service applications; educational applications, including computer-based learning and lesson applications and testing applications; screening applications; consumer preference monitoring applications; compliance applications, including applications that generate

10   notifications of compliance related activities, including notifications regarding product maintenance; test result applications, including applications that provide at least one of standardized tests results, consumer product test results, and maintenance results; and linking applications, including applications that link two or more of the above applications.

15   In the preferred embodiment, each speech recognition application includes an application file programmed into the speech recognition system 16. Preferably, the series of queries that make up the application is designed to obtain specific information from the respondents to aid in customer or consumer service, education and research and development of particular products or services or other functions. For example, a

20   particular speech application could be designed to ask respondents specific queries about a particular product or service. The entity that issues the application may then use this information to further develop the particular product or service. An application may also be used to provide specific information to a particular person or department.

     Fig. 2 is a flow diagram which shows the method of adapting a speech recognition

25   application and training a speech recognition application respondent in order to enable the respondent to effectively complete the application. First, either the automatic calling system 14 initiates a call to the target person at telephone 18, or the target person initiates a telephone call to the system 12 based on information provided to the respondent by the organization providing the application. The system 12 initiates the application by

providing an introduction to the respondent, stage 22. The introduction generally identifies the host organization and informs the respondent of the purpose of the application.

In stage 24, the system 12 provides a brief explanation of the application, including the fact that the respondent is speaking to a computer that is only capable of posing queries, recognizing certain of the respondent's responses The system then prompts the respondent to affirm that he or she understands how to interact with the system 12. This prompt enables the system 12 to determine if the respondent is capable of interacting with an automated speech recognition system. Based on the response given, the system determines which step will be executed next. If the respondent replies quickly with a "yes" or some similar affirmation, the system may move on to the identification check, stage 26, in which the respondent is asked to provide identification, typically in the form of a personal identification number (PIN), voice verification, or other method. While the use of a PIN is desirable in application surveys that address private matters concerning the respondent, the use of a PIN is not required in the present invention.

If the respondent answers "no" or does not respond to affirmation request in stage 24, the system 12 explains in greater detail how the system operates. The system prompts the respondent to answer "Hello" to a similar greeting offered by the system, as a training exercise for the respondent. If the respondent replies correctly, the system can repeat the explanation of the system and proceed to the identification stage 26. If the respondent is does not reply to the greeting request or replies with a reply that is not understood by the system 12, the system can initiate several more attempts at, and approaches to trying to explain the process to the respondent, including attempting to determine whether the respondent is having difficulty hearing the application, in which the system 12 would be instructed to increase the volume of the prompts and/or to slow the speed at which the prompts are played by the system 12. If the system is unable to teach the respondent how to respond to the application, the system enters an end call stage 25, in which the respondent is thanked and optionally informed that they will be contacted by a human being, and the call is terminated.

In optional identification stage 26, the respondent is asked for identification, which in one example may include a PIN. If the PIN is correctly input either by speaking the

numbers or by pressing the number on the telephone keypad, the application moves to the instruction step 28. If the respondent enters an incorrect PIN or does not know his or her PIN, the system enters an end call stage 25, in which the respondent is thanked and optionally informed how they can obtain a proper PIN, and the call is terminated.

5   After the identity of the respondent has been confirmed in step 26, the system enters instruction stage 28. In instruction stage 28, the system 12 explains the purpose of the application and the benefits provided by the application. The system 12 explains the structure of the application and informs the respondent of what types of answers are necessary for the application to be successful. The system 12 can then provide a sample

10   prompt to the respondent in order to prepare the respondent for what to expect during the actual application. If the survey includes a rating system, it is explained in this stage and the sample question can require an answer that uses the rating system. An example of this process in shown in Figs. 3A-3C, which include an example question and the options available, depending on the responses given. If, in this stage, the respondent is unable to

15   answer the sample prompt satisfactorily, the system enters an end call stage 25, in which the respondent is thanked and optionally informed that they will be contacted by a human being, and the call is terminated.

       After stage 28 has been completed satisfactorily, the system enters stage 30, in which the prompts of the application are presented to the respondent. At any point during

20   stage 30, if the respondent does not understand the process or becomes confused by the application, prompts or rating system, the system 12 can re-enter either or both of explanation stage 24 and instruction stage 28 to provide help for the respondent, as necessary. The system 12, when appropriate, can then return to survey stage 30 to complete the application. During the application, the system records each of the responses

25   provided by the respondent for review at a later time.

       At the completion of the application, the system enters a "wrap up" stage 32 in which the respondent is informed that the survey is over and is thanked by the host organization for participating in the application. Application feedback stage 34 provides an opportunity for the respondent to have his or her comments regarding the application

itself or regarding the speech recognition application system recorded for review by the host organization.

Accordingly, the present invention enables the system 12 both to train the respondent in properly responding to the prompts of the associated application and to alter

5   the course of the application based on responses to introductory and explanatory prompts. For example, if the respondent, from the beginning of the call, understands the application process and is capable of responding to the prompts, the explanation stage 24 and instruction stage 28 can be quickly navigated through, saving time and money for the host organization, since more respondents can be processed in a given period of time. On the

10   other hand, if the respondent is having difficulty understanding or hearing the system 12, the system is able to offer further explanations, training and sample prompts and, if the person is still not able to complete the survey, the system 12 is able to terminate the application.

The invention may be embodied in other specific forms without departing from the

15   spirit or essential characteristics thereof. The present embodiments are therefore to be considered in respects as illustrative and not restrictive, the scope of the invention being indicated by the appended claims rather than by the foregoing description, and all changes which come within the meaning and range of the equivalency of the claims are therefore intended to be embraced therein.